BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e34742
doi: 10.3897/biss.3.34742

OPEN ACCESS

Conference Abstract

Wikidata and the biodiversity knowledge graph

Roderic Page ‡

‡ University of Glasgow, Glasgow, United Kingdom

Corresponding author: Roderic Page (roderic.page@glasgow.ac.uk)

Received: 21 Mar 2019 | Published: 13 Jun 2019

Citation: Page R (2019) Wikidata and the biodiversity knowledge graph. Biodiversity Information Science and Standards 3: e34742. https://doi.org/10.3897/biss.3.34742

Abstract

This talk explores the role Wikidata (Vrandečić and Krötzsch 2014) might play in the task of assembling biodiversity information into a single, richly annotated and cross linked structure known as the biodiversity knowledge graph (Page 2016). Initially conceived as a language-independent data store of facts derived from the Wikipedia, Wikidata has morphed into a global knowledge graph, complete with a user friendly interface for data entry and a powerful implementation of the SPARQL query language. Wikidata already underpins projects such as Gene Wiki (Burgstaller-Muehlbacher et al. 2016) and Scholia (Nielsen et al. 2017). Much of the content of Wikispecies is being automatically added to Wikidata, hence many of the entities relevant to biodiversity (such as taxa, taxonomic publications, and taxonomists) well represented in Wikidata, making it even more attractive.

Much of the data relevant to biodiversity is widely scattered in different locations, requiring considerable manual effort to collect and curate. Appeals to the taxonomic community to undertake these tasks have not always met with success. For example, the Global Registry of Biodiversity Repositories (GrBio) was an attempt to create a global list of biodiversity repositories, such as natural history museums and herbaria. An appeal by Schindel et al. (2016) for the taxonomic community to curate this list largely fell on deaf ears, and at the time of writing the GrBio project is moribund. Given that many repositories are housed in institutions that are the subject of articles in Wikipedia, many of these repositories already have entries in Wikidata. Hence, rather than follow the route GrBio took of building a resource and then hoping a community will assemble around that resource, we could go to Wikidata where there is an existing community and build the resource there. An impressive

© Page R. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

example of the potential for this is WikiCite, which initially had the goal of including in Wikidata every article cited in any of the Wikipedias. Taxonomic articles are highly cited in Wikipedia (Nielsen 2007), hence already fall within the remit of WikiCite. Hence Wikidata is a candidate for the "bibliography of life" (King et al. 2011), a database of all taxonomic literature.

Another important role Wikidata can play is to define the boundaries of a biodiversity knowledge graph. Entities such as journals, articles, people, museums, and herbaria are often already in Wikidata, hence we can delegate managing that content to the Wikidata community (bolstered by our own contributions), and focus instead on domain-specific entities such as DNA sequences, specimens, etc., or domain specific attributes of those entities if they are already in Wikidata. This means we can avoid the inevitable "mission creep" that bedevils any attempt to link together information from multiple disciplines.

These ideas are explored using examples based on content entirely within Wikidata (including entities such as publications, authorship, and natural history collections), as well as approaches that combine Wikidata with external knowledge graphs such as Ozymandias (Page 2018).

Keywords

biodiversity knowledge graph, linked data, wikidata, wikicite, bibliography of life

Presenting author

Roderic Page

References

- Burgstaller-Muehlbacher S, Waagmeester A, Mitraka E, Turner J, Putman T, Leong J, Naik C, Pavlidis P, Schriml L, Good BM, Su AI (2016) Wikidata as a semantic framework for the Gene Wiki initiative. Database 2016: baw015-baw015. https://doi.org/10.1093/database/baw015
- King D, Morse D, Willis A, Dil A (2011) Towards the bibliography of life. ZooKeys 150: 151-166. https://doi.org/10.3897/zookeys.150.2167
- Nielsen FA (2007) Scientific citations in Wikipedia. arXiv:0705.2106 [cs].
- Nielsen FÅ, Mietchen D, Willighagen E (2017) Scholia, Scientometrics and Wikidata. Lecture Notes in Computer Science 237-259. https://doi.org/10.1007/978-3-319-70407-4_36
- Page RDM (2016) Towards a biodiversity knowledge graph. Research Ideas and Outcomes 2: e8767-e8767. https://doi.org/10.3897/rio.2.e8767
- Page RM (2018) Ozymandias: A biodiversity knowledge graph. bioRxiv 485854-485854. https://doi.org/10.1101/485854

- Schindel D, Miller S, Trizna M, Graham E, Crane A (2016) The Global Registry of Biodiversity Repositories: A Call for Community Curation. Biodiversity Data Journal 4: e10293-e10293. https://doi.org/10.3897/BDJ.4.e10293
- Vrandečić D, Krötzsch M (2014) Wikidata. Communications of the ACM 57 (10): 78-85. https://doi.org/10.1145/2629489